# Efficient solution of anisotropic lattice equations by the recovery method

Babuska, I ; Sauter, Stefan A

Abstract: In a recent paper, the authors introduced the recovery method (local energy matching principle) for solving large systems of lattice equations. The idea is to construct a partial differential equation along with a finite element discretization such that the arising system of linear equations has equivalent energy as the original system of lattice equations. Since a vast variety of efficient solvers is available for solving large systems of finite element discretizations of elliptic PDEs, these solvers may serve as preconditioners for the system of lattice equations. In this paper, we will focus on both the theoretical and the numerical dependence of the method on various mesh-dependent parameters, which can be easily computed and monitored during the solution process. Systematic parameter tests have been performed which underline (a) the robustness and the efficiency of the recovery method and (b) the reliability of the control parameters, which are computed in a preprocessing step to predict the performance of the preconditioner based on the recovery method.

DOI: https://doi.org/10.1137/070690717

# Efficient Solution of Anisotropic Lattice Equations by the Recovery Method.

I. Babuška* S.A. Sauter†

April 30, 2007

**Abstract**

In a recent paper, the authors introduced the *recovery method* resp. *local energy matching principle* for solving large systems of lattice equations. The idea is to construct a partial differential equation along with a finite element discretisation such that the arising system of linear equations has equivalent energy as the original system of lattice equations.

Since a vaste variety of efficient solvers is available for solving large systems of finite element discretisations of elliptic PDEs these solvers may serve as preconditioners for the system of lattice equations.

In this paper, we will focus on both, the theoretical and the numerical dependence of the method on various mesh-dependent parameters which can be easily computed and monitored during the solution process. Systematic parameter tests have been performed which underline (a) the robustness and the efficiency of the recovery method and (b) the reliability of the control parameters which are computed in a preprocessing step to predict the performance of the preconditioner based on the recovery method.

## 1 Introduction

Lattice models are used in many applications such as models of heterogeneous materials ([18], [11]), fracture models ([19]), porous media ([9], [8]), biophysics ([14]), and chip design. For a survey of some applications, we refer to [18] and [20]. Lattices are becoming more and more interesting for industrial production because these materials are light, cheap, and can be designed to prescribed stiffness requirements. From the viewpoint of numerical modelling, such problems are challenging because

*ICES, University of Texas at Austin, Austin, TX 78712, USA

†(stas@math.unizh.ch), Institut für Mathematik, Universität Zürich, Winterthurerstr 190, CH-8057 Zürich, Switzerland

the geometry of the lattice, typically, is very complicated and consists of a huge number of rods or beams. Hence, the efficient numerical solution of the arising systems of linear equation is non-trivial because efficient solvers such as, e.g., multigrid methods, cannot be applied in a straightforward way. The reason is that the equations are not formulated on an Euclidean domain or a hyperplane and, hence, a grid hierarchy is **not** available.

Our paper deals we two types of problems in this field: (a) the construction of an efficient preconditioner for some lattice equation and (b) the construction of an elliptic PDE with *homogenized* discontinuous coefficients on an Euclidean domain such that the finite element discretization thereof leads to a linear system of equation which, locally, is spectrally equivalent to the original lattice equation. This step paves the way to apply some numerical upscaling techniques – applied to the "recovered" elliptic differential equation on an Euclidean domain – (cf. [1], [12], [15], [16], [6], [22]) in order to homogenize these coefficients to even coarser scales.

We emphasize that our approach is by no means related to a period setting but can be applied to general lattices. During the computation, some constants are determined which will serve as indicators for the efficiency of our method and guarantee that the algorithm does not fail in an unpredictable way. In this paper, which can be regarded as Part II of [2], we will focus on the algorithmic formulation of the recovery method and systematic numerical parameter tests.

The main results of this paper are: a) The algorithm is very robust and works also for complex applications such as the electrostatic problem in a *routing channel*. This problem has an engineering importance and leads to a highly anisotropic lattice of a very extreme character. b) The theoretical indicators for the performance of the method predicts very well the true performance which was observed numerically. This was especially addressed in the case of the routing channel. Also here, the theoretical indicators well predicted the need of a larger number of iteration for solving the arising system of linear equations.

A related paper, where the preconditioning of elastic problems on periodic structures has been investigated, is [25]. Another class of efficient solvers for such problems are *algebraic* multigrid methods (cf. [17], [5], [21], [28], [3]). Some standard references to multigrid methods are [13], [4], [26], [27].

# 2 Model Problem

## 2.1 Setting

Let $\Theta := \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$ denote the set of nodal points and let $\mathcal{E} \subset \Theta \times \Theta$ be a symmetric set of edges, i.e., $e = (x, y) \in \mathcal{E}$ implies $(y, x) \in \mathcal{E}$. The set of nodal points together with the set of edges $\mathcal{E}$ form the graph $\mathcal{G}$ of the lattice.

From the physical point of view, we shall deal with equations on the lattice $\mathcal{G}$ which are of the same (abstract) form as the equations of linear electrostatics on

complicated electric circuits, i.e., are described by scalar discrete potential equations of second order. First, we will consider the case that no essential constraints at the nodes are described. The case of essential constraints will be treated in Section 5.

The electric conductivity through an edge $(x, y) \in \mathcal{E}$ is described by a symmetric, positive mapping $\mathbf{a} = (a_e)_{e \in \mathcal{E}}$ with

$$\left.\begin{array}{c} a_{(x,y)} = a_{(y,x)} \\ a_{(x,y)} > 0 \end{array}\right\} \qquad \forall\, (x, y) \in \mathcal{E}.$$

Let $S$ denote the space of (unconstrained) grid functions

$$S := \mathbb{R}^{\Theta} := \{\mathbf{u} \mid \mathbf{u} : \Theta \to \mathbb{R}\}. \tag{2.1}$$

On $S$, we introduce the bilinear form

$$\mathbf{B}(\mathbf{u}, \mathbf{v}) := \frac{1}{2} \sum_{e=(x,y) \in \mathcal{E}} \frac{a_e}{h_e} (u_y - u_x)(v_y - v_x), \tag{2.2}$$

where $h_e := \|x - y\|$. We introduce the quotient space $V := S/\mathbb{R}$ where the equivalence classes are formed by functions which differ only by a constant grid function. We consider the following Poisson-type problem:

Let $F \in V'$ be given. Find $\mathbf{u} = (u_x)_{x \in \Theta} \in V$ so that

$$\mathbf{B}(\mathbf{u}, \mathbf{v}) = F(\mathbf{v}) \qquad \forall \mathbf{v} = (v_x)_{x \in \Theta} \in V. \tag{2.3}$$

This equation has a unique solution as can be seen from the following well-known theorem.

**Theorem 2.1** *Let the lattice be connected. Then,*

*a.* $\mathbf{B}(\cdot, \cdot) : V \times V \to \mathbb{R}$ *is a scalar product and* $\mathbf{B}(\mathbf{u}, \mathbf{u})^{1/2}$ *a norm on* $V$*,*

*b. the variational problem (2.3) has a unique solution* $u \in V$ *for any right-hand side* $F \in V'$*.*

The variational problem (2.3) can be interpreted as a system of finite difference equations: We are testing equation (2.3) for all $z \in \Theta$ with the unit vectors $\mathbf{e}_z = (e_{z,x})_{x \in \Theta} \in S$, where

$$e_{z,x} := \begin{cases} 1 & x = z, \\ 0 & x \in \Theta \backslash \{z\}. \end{cases}$$

For $z \in \Theta$, we obtain the relation

$$\frac{1}{2} \sum_{e=(x,y) \in \mathcal{E}} \frac{a_e}{h_e} (u_y - u_x)(e_{z,y} - e_{z,x}) = \sum_{\substack{x \in \Theta: \\ e=(z,x) \in \mathcal{E}}} \frac{a_e}{h_e} (u_z - u_x).$$

By setting $F_z := F(\mathbf{e}_z)$ and

$$A_{xy} := \begin{cases} \displaystyle\sum_{\substack{z\in\Theta:\\ e=(x,z)\in\mathcal{E}}} a_e/h_e & \text{if } x = y, \\ -a_e/h_e & \text{if } e = (x,y) \in \mathcal{E}, \\ 0 & \text{otherwise,} \end{cases}$$

we obtain the finite difference equations

$$\sum_{y\in\Theta} A_{xy} u_y = F_x \qquad \forall x \in \Theta,$$

and use the short notation $\mathbf{Au} = \mathbf{F}$. To get an equivalent system to the variational formulation (2.3), we have to restrict the right-hand side and the solution in (2.4) to appropriate quotient spaces: For given $\mathbf{F} \in V'$, find $\mathbf{u} \in V$ such that

$$\mathbf{Au} = \mathbf{F}. \tag{2.4}$$

# 3 The Recovery Method

The recovery method for transferring given lattice equations into a continuous partial differential equation for which efficient solvers are available has been introduced in [2]. These efficient solvers then may serve, via the recovery method, as a preconditioner for the given lattice equation. In this paper, we present an improved recovery strategy which allows to treat lattice equations with strong anisotropies in the coefficients in a robust way.

As in the previous method, the construction consists of the definition of a domain $\Omega \subset \mathbb{R}^d$ for the continuous problem and the definition of the coefficient function in the partial differential equation.

We begin with the definition of the domain $\Omega$. For a subset $M \subset \mathbb{R}^d$, we write $\operatorname{int}(M)$ for the interior of $M$.

**Theorem 3.1** *Let $\Theta \subset \mathbb{R}^d$, $d \in \{2,3\}$, denote a discrete set of points with $\operatorname{card}\Theta \geq d+1$. Then, the Voronoï method defines a simplicial mesh $\mathcal{G}_{\mathrm{FE}}$ of $d$-dimensional, disjoint simplices where the set of mesh points $\Theta_{\mathrm{FE}}$ satisfies $\Theta_{\mathrm{FE}} = \Theta$. For non-identical elements $\tau, t \in \mathcal{G}_{\mathrm{FE}}$, the intersection $\overline{\tau} \cap \overline{t}$ is either empty, a common point, a common edge, or – for $d = 3$ – a common face.*

The mesh $\mathcal{G}_{\mathrm{FE}}$ covers the set

$$\Omega := \operatorname{int} \overline{\bigcup_{\tau\in\mathcal{G}_{\mathrm{FE}}} \tau}.$$

**Remark 3.2** *An algorithm for assembling a triangulation (Delaunay triangulation) as in Theorem 3.1 is described, e.g., in [10], [23], [24], [7].*

**Assumption 3.3** *The set $\Omega \subset \mathbb{R}^d$ is a polygonal (polyhedral for $d = 3$) Lipschitz domain.*

The existence of the triangulation $\mathcal{G}_{\mathrm{FE}}$ does **not** ensure that the parameters which are measuring the quality of the triangles, e.g., the maximal/minimal angle or the maximal ratio of diameters of neighbouring elements, is moderately bounded (in fact, if all nodal points lie, e.g., on a straight line all triangles in $\mathcal{G}_{\mathrm{FE}}$ are degenerate to a line). In this light, we will introduce some mesh-dependent parameters which may serve as indicators for the performance of the recovery method.

**Definition 3.4** *The shape regularity of the mesh $\mathcal{G}_{\mathrm{FE}}$ is characterized by*

$$C_{sr} := \max_{\tau \in \mathcal{G}_{\mathrm{FE}}} \frac{h_\tau}{\rho_\tau}, \tag{3.1}$$

*where $h_\tau := \operatorname{diam} \tau$ and $\rho_\tau$ is the radius of the largest inscribed ball in $\tau$.*

We make an assumption on the "compatibility" of the meshes and introduce some notation.

Let the edges in $\mathcal{G}_{\mathrm{FE}}$ be denoted by $\mathcal{E}_{\mathrm{FE}}$. To distinguish in the notation the edges in $\mathcal{E}_{\mathrm{FE}}$ from edges in the given lattice $\mathcal{E}$ we will use a tilde superscript for edges in $\mathcal{E}_{\mathrm{FE}}$. For $\tilde{e} = (x, y) \in \mathcal{E}_{\mathrm{FE}}$, we have $x, y \in \Theta$ and we may associate with $\tilde{e}$ a path $\pi(\tilde{e}) = \left(e_1, e_2, \ldots, e_{q(\tilde{e})}\right) \subset \mathcal{E}$ such that

$$x_0 = x, \quad x_{q(\tilde{e})} = y \quad \text{and} \quad e_i = (x_{i-1}, x_i), \qquad 1 \leq i \leq q(\tilde{e})$$

connecting $x$ and $y$. In an analogous way, we associate such a connecting path $\pi_{\mathrm{FE}}(e) \subset \mathcal{E}_{\mathrm{FE}}$ for each $e \in \mathcal{E}$. In general, the paths $\pi_{\mathrm{FE}}(e)$ and $\pi(\tilde{e})$ are by no means uniquely determined. In Section 3.1, we will derive a selection criterion for a proper choice of theses paths. A minimal requirement is that, in the case $e \in \mathcal{E}_{\mathrm{FE}} \cap \mathcal{E}$, we choose

$$\pi_{\mathrm{FE}}(e) = \pi(e) = (e).$$

**Assumption 3.5** *The lattice $\mathcal{G}$ and the mesh $\mathcal{G}_{\mathrm{FE}}$ are connected.*

**Remark 3.6** *The connectivity of the lattice $\mathcal{G}$ and the connectivity of the mesh $\mathcal{G}_{\mathrm{FE}}$ imply that $\pi(\tilde{e}) \neq \emptyset$ for every $\tilde{e} \in \mathcal{E}_{\mathrm{FE}}$ and $\pi_{\mathrm{FE}}(e) \neq \emptyset$ for every $e \in \mathcal{E}$.*

Our goal is to replace the lattice equations (2.3) by a finite element discretisation of a Poisson equation on the mesh $\mathcal{G}_{\mathrm{FE}}$. This is done in two steps.

(a) Define a system of lattice equations on the edges $\mathcal{E}_{\mathrm{FE}}$ of $\mathcal{G}_{\mathrm{FE}}$ which has equivalent energy.

(b) Replace the lattice equations on $\mathcal{E}_{\mathrm{FE}}$ by an averaged (possibly anisotropic) Poisson problem on $\Omega$.

## 3.1 Definition of a System of Lattice Equations on $\mathcal{E}_{\mathrm{FE}}$ with Equivalent Energy

In this section, we will introduce a system of lattice equations on the set of finite element edges which has equivalent energy as the original equations. Our approach is based on a suitable local average of the conductivity coefficients $(a_e)_{e\in\mathcal{E}}$ along the paths $\pi(\tilde{e})$. In this light, we will introduce some notations.

**Notation 3.7** *For $\tilde{e} = (x, y) \in \mathcal{E}_{\mathrm{FE}}$, let*

$$a_{\tilde{e}}^{\mathrm{FE}} := h_{\tilde{e}} / \left( \sum_{e\in\pi(\tilde{e})} \frac{h_e}{a_e} \right). \tag{3.2}$$

*For an edge $e \in \mathcal{E}$, we define*

$$\delta_e := \sum_{\tilde{e}\in\pi_{\mathrm{FE}}(e)} \frac{a_e}{h_e} / \frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}} \quad \textit{and} \quad \delta_{\max} := \max_{\tilde{e}\in\mathcal{E}_{\mathrm{FE}}} \sum_{e\in\mathcal{E}:\tilde{e}\in\pi_{\mathrm{FE}}(e)} \delta_e. \tag{3.3}$$

We introduce the bilinear form

$$\mathbf{B}_{\mathrm{FE}}(\mathbf{u}, \mathbf{v}) := \frac{1}{2} \sum_{(x,y)\in\mathcal{E}_{\mathrm{FE}}} a_{(x,y)}^{\mathrm{FE}} \frac{(u_y - u_x)(v_y - v_x)}{\|x - y\|}. \tag{3.4}$$

**Theorem 3.8** *Let Assumptions 3.5 be satisfied. Then, the estimate*

$$\mathbf{B}(\mathbf{u}, \mathbf{u}) \le \delta_{\max} \mathbf{B}_{\mathrm{FE}}(\mathbf{u}, \mathbf{u}) \qquad \forall \mathbf{u} \in \mathbb{R}^{\Theta} \tag{3.5}$$

*holds with $\delta_{\max}$ as in (3.3).*

**Proof.** For $e = (x, y) \in \mathcal{E}$, we get

$$\begin{aligned} \frac{a_e}{h_e}(u_x - u_y)^2 &= \left( \sum_{\tilde{e}=(z_1,z_2)\in\pi_{\mathrm{FE}}(e)} \sqrt{\frac{a_e}{h_e} / \frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}}} \sqrt{\frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}}} (u_{z_2} - u_{z_1}) \right)^2 \\ &\le \delta_e \sum_{\tilde{e}=(z_1,z_2)\in\pi_{\mathrm{FE}}(e)} \frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}} (u_{z_2} - u_{z_1})^2 \end{aligned}$$

and, hence,

$$\mathbf{B}(\mathbf{u}, \mathbf{u}) = \frac{1}{2} \sum_{e=(x,y)\in\mathcal{E}} \frac{a_e}{h_e} (u_y - u_x)^2 \le \frac{1}{2} \sum_{e\in\mathcal{E}} \delta_e \sum_{\tilde{e}=(z_1,z_2)\in\pi_{\mathrm{FE}}(e)} \frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}} (u_{z_2} - u_{z_1})^2 \tag{3.6}$$

$$= \frac{1}{2} \sum_{\tilde{e}=(x,y)\in\mathcal{E}_{\mathrm{FE}}} \frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}} (u_x - u_y)^2 \sum_{e\in\mathcal{E}:\tilde{e}\in\pi(e)} \delta_e \tag{3.7}$$

$$\le \delta_{\max} \mathbf{B}_{\mathrm{FE}}(\mathbf{u}, \mathbf{u}).$$

■

The recovery method for the lattice equations will employ two-sided estimates in the energy bilinear forms $\mathbf{B}$ and $\mathbf{B}_{\mathrm{FE}}$. In this light, we will consider next the opposite estimate.

**Theorem 3.9** *Let Assumptions 3.5 be satisfied. The bilinear forms* $\mathbf{B}_{\mathrm{FE}}$ *and* $\mathbf{B}$ *satisfy*

$$\frac{1}{\delta_{\max}}\mathbf{B}(\mathbf{u},\mathbf{u}) \leq \mathbf{B}_{\mathrm{FE}}(\mathbf{u},\mathbf{u}) \leq \overline{n}\mathbf{B}(\mathbf{u},\mathbf{u}) \qquad \forall \mathbf{u} \in \mathbb{R}^{\Theta}$$

*with*

$$\overline{n} := \max_{e\in\mathcal{E}} \operatorname{card}\left\{\tilde{e} \in \mathcal{E}_{\mathrm{FE}} : e \in \pi(\tilde{e})\right\}.$$

**Proof.** The left inequality is (3.5) and, hence, we consider here only the right one.

Note that the definition of the averaged coefficients $a_{\tilde{e}}^{\mathrm{FE}}$ in (3.2) implies

$$\sum_{e\in\pi(\tilde{e})} \frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}} \Big/ \frac{a_e}{h_e} = 1.$$

Hence, for $\tilde{e} = (x,y) \in \mathcal{E}_{\mathrm{FE}}$, we get

$$\begin{aligned}\frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}}(u_x-u_y)^2 &= \left(\sum_{e=(z_1,z_2)\in\pi(\tilde{e})} \sqrt{\frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}}\Big/\frac{a_e}{h_e}}\sqrt{\frac{a_e}{h_e}}(u_{z_2}-u_{z_1})\right)^2 \\ &\leq \sum_{e=(z_1,z_2)\in\pi(\tilde{e})} \frac{a_e}{h_e}(u_{z_2}-u_{z_1})^2.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbf{B}_{\mathrm{FE}}(\mathbf{u},\mathbf{u}) &= \frac{1}{2}\sum_{\tilde{e}=(x,y)\in\mathcal{E}_{\mathrm{FE}}} \frac{a_{\tilde{e}}}{h_{\tilde{e}}}(u_y-u_x)^2 \leq \frac{1}{2}\sum_{\tilde{e}\in\mathcal{E}_{\mathrm{FE}}}\sum_{e=(z_1,z_2)\in\pi(\tilde{e})} \frac{a_e}{h_e}(u_{z_2}-u_{z_1})^2 \\ &= \frac{1}{2}\sum_{e=(x,y)\in\mathcal{E}} \frac{a_e}{h_e}(u_x-u_y)^2 \operatorname{card}\left\{\tilde{e}\in\mathcal{E}_{\mathrm{FE}} : e\in\pi(\tilde{e})\right\} \\ &\leq \overline{n}\mathbf{B}(\mathbf{u},\mathbf{u}).\end{aligned}$$

■

**Remark 3.10** *(a) In the special case* $\mathcal{E}_{\mathrm{FE}} = \mathcal{E}$*, we choose* $\pi(e) = \pi_{\mathrm{FE}}(e) = e$*. Hence, the bilinear forms* $\mathbf{B}_{\mathrm{FE}}$ *and* $\mathbf{B}$ *coincide (cf. (3.2), (3.3)).*

*(b) Note that the constants* $\delta_e$ *in (3.5) are moderately bounded also for conductivity coefficients with large global ratio* $\max_{e\in\mathcal{E}} a_e / \min_{e\in\mathcal{E}} a_e$ *as long as the local variations (measured by* $\max_{\tilde{e}\in\mathcal{E}_{\mathrm{FE}}:e\in\pi(\tilde{e})} \frac{a_e}{h_e}\Big/\frac{a_{\tilde{e}}^{\mathrm{FE}}}{h_{\tilde{e}}}$*) are moderately bounded.*

*(c) If the lattice equation arise from a finite element discretisation of a Poisson-type problem on a triangulation of the domain, we may choose the mesh $\mathcal{G}_{\mathrm{FE}}$ as the original finite element mesh and the constants in the energy estimate equal 1. This is independent of possibly large jumps in the coefficients of the original problem.*

In an ideal situation, the paths $\pi(\tilde{e})$, $\pi_{\mathrm{FE}}(e)$ should be chosen such that $\delta_{\max}$ is minimal. Since this (global) optimization would be too time consuming we propose to select the paths such that the quantities $\delta_e$ are small. We have

$$\max_{e\in\mathcal{E}} \delta_e = \max_{e\in\mathcal{E}} \sum_{\tilde{e}\in\pi_{\mathrm{FE}}(e)} \frac{a_e}{h_e} \frac{h_{\tilde{e}}}{a_{\tilde{e}}^{\mathrm{FE}}}.$$

Since the coefficients $a_{\tilde{e}}^{\mathrm{FE}}$ depends on the selection of the paths $\pi(e)$ (cf. (3.2)) we obtain

$$\max_{e\in\mathcal{E}} \delta_e = \max_{e\in\mathcal{E}} \frac{a_e}{h_e} \sum_{\tilde{e}\in\pi_{\mathrm{FE}}(e)} \sum_{e'\in\pi(\tilde{e})} \frac{h_{e'}}{a_{e'}}.$$

This leads to the strategy: First, choose the paths $\pi(\tilde{e})$ such that

$$\gamma_{\tilde{e}} := \sum\nolimits_{e'\in\pi(\tilde{e})} h_{e'}/a_{e'}$$

is small and then choose the paths $\pi_{\mathrm{FE}}(e)$ such that $\delta_e$ is small.

**Algorithm 3.11 (Selection of $\pi(e)$, $\pi_{\mathrm{FE}}(e)$)** *1. For any $\tilde{e} \in \mathcal{E}_{\mathrm{FE}}$, determine $\pi(\tilde{e})$ via the condition*

$$\gamma_{\tilde{e}} = \sum_{e\in\pi(\tilde{e})} h_e/a_e = \min_{\pi\in\mathfrak{P}(\tilde{e})} \sum_{e\in\pi} h_e/a_e, \tag{3.8}$$

*where $\mathfrak{P}(\tilde{e})$ denotes the set of all paths in $\mathcal{E}$ connecting the endpoints of $\tilde{e}$.*

*2. For any $e \in \mathcal{E}$, determine $\pi_{\mathrm{FE}}(e)$ via the condition*

$$\sum_{\tilde{e}\in\pi_{\mathrm{FE}}(e)} \gamma_{\tilde{e}} = \min_{\pi_{\mathrm{FE}}\in\mathfrak{P}_{\mathrm{FE}}(e)} \sum_{\tilde{e}\in\pi_{\mathrm{FE}}} \gamma_{\tilde{e}}. \tag{3.9}$$

In order to keep the minimisation process in Algorithm 3.11 local, we recommend to locally search for minimal paths in recursively defined layers of edges about the given edge $\tilde{e}$ (in (3.8)) resp. $e$ in (3.9). We skip the detailed formulation of this recursive procedure.

## 3.2 Recovery of the Continuous Variational Form

In this step, we will define, for the given system of lattice equations, a bilinear form on the continuous level along with a transfer mapping which has equivalent energy. For the continuous problem, we employ as an ansatz an anisotropic Poisson problem of the form

$$a(u,v) := \sum_{\tau\in\mathcal{G}_{\mathrm{FE}}} \int_\tau \langle \nabla v, \mathbf{A}_\tau \nabla u\rangle ,$$

where the diffusion matrix $\mathbf{A}_\tau$ is constant on each simplex $\tau$

$$\mathbf{A}_\tau = \left(a_\tau^{ij}\right)_{i,j=1}^{d} \quad \text{with} \quad a_\tau^{ij} = a_\tau^{ji} \quad \forall 1 \le i,j \le d.$$

For the construction of $\mathbf{A}_\tau$, we start with some preliminaries on local finite element matrices and associated finite difference operators. Consider a simplex $\tau = \operatorname{conv}\{\mathbf{x}_1,\dots,\mathbf{x}_{d+1}\} \in \mathcal{G}_{\mathrm{FE}}$ and denote by $b_i$, $1 \le i \le d+1$, the corresponding local affine Lagrange basis ("hat functions") on $\tau$. The local finite element stiffness matrix $\mathbf{L}_\tau = (L_{i,j})_{i,j=1}^{d+1}$ for the bilinear form $a(\cdot,\cdot)$ is defined by

$$L_{i,j} := \int_\tau \langle \nabla b_i, \mathbf{A}_\tau \nabla b_j\rangle \, d\mathbf{x} \qquad 1 \le i,j \le d+1.$$

Let $\mathbf{u} = (u_i)_{i=1}^{d+1} \in \mathbb{R}^{d+1}$ be a grid function with values $u_i$ at $\mathbf{x}_i$, $1 \le i \le d+1$. For simplicity, we set $\mathbf{x}_{d+2} := \mathbf{x}_1$ and $\mathbf{x}_0 := \mathbf{x}_{d+1}$ and use this convention also for $\mathbf{u}$ and $\mathbf{v}$. The set of edges for a simplex $\tau$ is denoted by $\mathcal{E}_\tau$ and the set of vertices by $\mathcal{V}_\tau$. Let

$$\mathbf{e}_i := \mathbf{x}_{i+1} - \mathbf{x}_{i-1}, \qquad |\tau| := \operatorname{volume}(\tau). \tag{3.10}$$

For the remaining part of this section, we restrict to the case $d = 2$. The formulae for the higher-dimensional case $d > 2$ can be derived in the same fashion.

For each edge $\tilde{e} \in \mathcal{E}_{\mathrm{FE}}$ we define the number of adjacent triangles by

$$t_e := \sharp\{\tau \in \mathcal{G}_{\mathrm{FE}} : \tilde{e} \subset \overline{\tau}\}$$

and for $e_i \in \mathcal{E}_\tau$ we write short $t_i$ for $t_{e_i}$.

**Lemma 3.12** *Let $d = 2$ and let the coefficients $\widehat{\alpha}_\tau^{11}$, $\widehat{\alpha}_\tau^{12} = \widehat{\alpha}_\tau^{21}$, $\widehat{\alpha}_\tau^{22}$ in the symmetric $2\times 2$ matrix $\widehat{\mathbf{A}}_\tau$ be defined by*

$$\begin{pmatrix} \widehat{\alpha}_\tau^{11} \\ \widehat{\alpha}_\tau^{12} \\ \widehat{\alpha}_\tau^{22} \end{pmatrix} := \frac{1}{|\tau|} \begin{bmatrix} e_{21}^2 & 2e_{21}e_{31} & e_{31}^2 \\ e_{21}e_{22} & e_{21}e_{32}+e_{22}e_{31} & e_{31}e_{32} \\ e_{22}^2 & 2e_{22}e_{32} & e_{32}^2 \end{bmatrix} \begin{pmatrix} \frac{\alpha_2}{h_2} + \frac{\alpha_1}{h_1} \\ \frac{\alpha_1}{h_1} \\ \frac{\alpha_3}{h_3} + \frac{\alpha_1}{h_1} \end{pmatrix}. \tag{3.11}$$

*Then,*

$$\mathbf{u}^{\intercal}\mathbf{L}_\tau\mathbf{u} = \sum_{e=(x,y)\in\mathcal{E}_\tau} \frac{\alpha_e}{h_e}\left(u_{i+1}-u_{i-1}\right)^2. \tag{3.12}$$

**Proof.** Let $\tau \in \mathcal{G}_{\mathrm{FE}}$ and choose a counterclockwise numbering of the vertices of $\tau$. For $\mathbf{v} \in \mathbb{R}^2$, let $\mathbf{v}^\perp := (v_2, -v_1)^\intercal$. By using (3.10) we obtain the representations

$$\nabla b_i = \frac{\mathbf{e}_i^\perp}{2\,|\tau|} \quad \text{and} \quad \mathbf{L}_\tau = \frac{1}{4\,|\tau|}\left[\left(\left\langle \mathbf{e}_i^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_j^\perp \right\rangle\right)_{i,j=1}^{3}\right].$$

Hence,

$$\begin{aligned}
\mathbf{u}^\intercal \mathbf{L}_\tau \mathbf{v} &= \sum_{i=1}^{3} u_i \sum_{j=1}^{3} \frac{\left\langle \mathbf{e}_i^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_j^\perp \right\rangle}{4\,|\tau|} v_j = \frac{1}{4\,|\tau|}\left\langle \sum_{i=1}^{3} u_i \mathbf{e}_i^\perp, \widehat{\mathbf{A}}_\tau \sum_{j=1}^{3} v_j \mathbf{e}_j^\perp \right\rangle \\
&= \frac{1}{4\,|\tau|}\left\langle (u_3-u_1)\,\mathbf{e}_3^\perp + (u_2-u_1)\,\mathbf{e}_2^\perp, \widehat{\mathbf{A}}_\tau \left((v_3-v_1)\,\mathbf{e}_3^\perp + (v_2-v_1)\,\mathbf{e}_2^\perp\right)\right\rangle \\
&= \frac{1}{4\,|\tau|}\begin{pmatrix} u_1-u_3 \\ u_2-u_1 \end{pmatrix}^\intercal \begin{bmatrix} \left\langle \mathbf{e}_3^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_3^\perp \right\rangle & -\left\langle \mathbf{e}_3^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_2^\perp \right\rangle \\ -\left\langle \mathbf{e}_3^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_2^\perp \right\rangle & \left\langle \mathbf{e}_2^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_2^\perp \right\rangle \end{bmatrix} \begin{pmatrix} v_1-v_3 \\ v_2-v_1 \end{pmatrix}.
\end{aligned}$$

On the other hand, we obtain

$$\begin{aligned}
\sum_{e=(x,y)\in\mathcal{E}_\tau} \frac{\alpha_e}{h_e}\,(u_{i+1}-u_{i-1})^2 &= \frac{\alpha_1}{h_1}\,(u_2-u_3)^2 + \frac{\alpha_2}{h_2}\,(u_3-u_1)^2 + \frac{\alpha_3}{h_3}\,(u_1-u_2)^2 \\
&= \begin{pmatrix} u_1-u_3 \\ u_2-u_1 \end{pmatrix}^\intercal \begin{bmatrix} \frac{\alpha_2}{h_2}+\frac{\alpha_1}{h_1} & \frac{\alpha_1}{h_1} \\ \frac{\alpha_1}{h_1} & \frac{\alpha_3}{h_3}+\frac{\alpha_1}{h_1} \end{bmatrix} \begin{pmatrix} u_1-u_3 \\ u_2-u_1 \end{pmatrix}^\intercal.
\end{aligned}$$

Hence, we have to choose the coefficients $\widehat{\alpha}_\tau^{11}$, $\widehat{\alpha}_\tau^{12}$, $\widehat{\alpha}_\tau^{22}$ in $\widehat{\mathbf{A}}_\tau$ such that

$$\frac{1}{4\,|\tau|}\left\langle \mathbf{e}_3^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_3^\perp \right\rangle = \frac{\alpha_2}{h_2}+\frac{\alpha_1}{h_1}, \quad \frac{1}{4\,|\tau|}\left\langle \mathbf{e}_2^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_2^\perp \right\rangle = \frac{\alpha_3}{h_3}+\frac{\alpha_1}{h_1}, \quad -\frac{1}{4\,|\tau|}\left\langle \mathbf{e}_3^\perp, \widehat{\mathbf{A}}_\tau \mathbf{e}_2^\perp \right\rangle = \frac{\alpha_1}{h_1}.$$

This linear system can be solved explicitly yielding the representation (3.11) for the coefficients in $\widehat{\mathbf{A}}_\tau$. ■

With these notations at hand, we can define the diffusion matrix by the following algorithm.

**Algorithm 3.13 (Computation of the element diffusion matrices for $d = 2$.)**

***For all*** $\tau \in \mathcal{G}_{\mathrm{FE}}$ ***do begin***

***compute***

$$\begin{pmatrix} \alpha_\tau^{11} \\ \alpha_\tau^{12} \\ \alpha_\tau^{22} \end{pmatrix} := \frac{1}{|\tau|}\begin{bmatrix} e_{21}^2 & 2e_{21}e_{31} & e_{31}^2 \\ e_{21}e_{22} & e_{21}e_{32}+e_{22}e_{31} & e_{31}e_{32} \\ e_{22}^2 & 2e_{22}e_{32} & e_{32}^2 \end{bmatrix} \begin{pmatrix} \frac{\alpha_2}{2h_2t_2}+\frac{\alpha_1}{2h_1t_1} \\ \frac{\alpha_1}{2h_1t_1} \\ \frac{\alpha_3}{2h_3t_3}+\frac{\alpha_1}{2h_1t_1} \end{pmatrix} \tag{3.13}$$

***assign***

$$\mathbf{A}_\tau := \begin{bmatrix} \alpha_\tau^{11} & \alpha_\tau^{12} \\ \alpha_\tau^{12} & \alpha_\tau^{22} \end{bmatrix}; \tag{3.14}$$

***end;***

The diffusion matrix $\mathbf{A}_\tau$ allows us to define the local bilinear form

$$L^\tau(u,v) := \int_\tau \langle \nabla u, \mathbf{A}_\tau \nabla v \rangle \, dx \qquad \forall u, v \in \mathbb{P}_1.$$

(Note that the scaling by $1/2$ in the right-hand side of (3.13) stems from the factor $1/2$ in (2.2).)

By summing over all local bilinear forms $L^\tau$ we derive the global variational formulation as follows.

Define the coefficient function $\mathbf{A} : \Omega \to \mathbb{R}^{2\times 2}$ by

$$\mathbf{A}|_\tau := \mathbf{A}_\tau. \tag{3.15}$$

The space of continuous piecewise linear functions on $\mathcal{G}_{\mathrm{FE}}$ is denoted by

$$S^{\mathrm{FE}} := \left\{ u \in C^0\left(\overline{\Omega}\right) \mid \forall \tau \in \mathcal{G}_{\mathrm{FE}} : u|_\tau \in \mathbb{P}_1 \right\}. \tag{3.16}$$

The standard local nodal basis is denoted by $(b_x)_{x\in\Theta}$. The finite element interpolation operator on $\mathcal{G}_{\mathrm{FE}}$ is denoted by by $I^{int}_{\mathrm{FE}} : \mathbb{R}^\Theta \to S^{\mathrm{FE}}$:

$$\left(I^{int}_{\mathrm{FE}}\mathbf{u}\right)(x) = \sum_{y\in\Theta} u_y b_y(x).$$

For $u, v \in S^{\mathrm{FE}}$, the global bilinear form which is associated to the lattice equations (3.4) is defined by

$$B_{\mathrm{FE}}(u,v) := \int_\Omega \langle \nabla u, \mathbf{A}\nabla v \rangle \, dx. \tag{3.17}$$

We will prove that this bilinear form has the same energy as the lattice equations on $\mathcal{E}_{\mathrm{FE}}$.

**Theorem 3.14** *Let Assumptions 3.3 and 3.5 be satisfied. For all* $\mathbf{u} \in \mathbb{R}^\Theta$ *and* $u := I^{int}_{\mathrm{FE}}\mathbf{u}$ *we have*

$$B_{\mathrm{FE}}(u,u) = \mathbf{B}_{\mathrm{FE}}(\mathbf{u},\mathbf{u}).$$

**Proof.** The result follows directly from Lemma 3.12 by summing over all triangles. ■

**Theorem 3.15** *Let Assumptions 3.3 and 3.5 be satisfied. For all* $\mathbf{u} \in \mathbb{R}^\Theta$ *and* $u := I^{int}_{\mathrm{FE}}\mathbf{u}$ *we have*

$$\frac{1}{\delta_{\max}}\mathbf{B}(\mathbf{u},\mathbf{u}) \le B_{\mathrm{FE}}(u,u) \le \overline{n}\mathbf{B}(\mathbf{u},\mathbf{u}).$$

# 4 A Finite Element Preconditioner based on the Recovery Method

In the previous section, we have introduced the recovery method which associates a variational formulation on the continuous level to the given lattice equations along a transfer mapping between discrete grid functions and finite element functions.

In this section, we will show that the stiffness matrix $\mathbf{A}_{\mathrm{FE}}$ which corresponds to the finite element discretisation of $B_{\mathrm{FE}}(\cdot,\cdot)$ on the mesh $\mathcal{G}_{\mathrm{FE}}$ is a quasi-optimal preconditioner of the system of lattice equations (2.4).

The preconditioned system takes the form

$$\mathbf{A}_{\mathrm{FE}}^{-1}\mathbf{A}\mathbf{u} = \mathbf{A}_{\mathrm{FE}}^{-1}\mathbf{F}.$$

Recall that the (restricted) matrix $\mathbf{A} : V \to V'$ (cf. (2.2)) is regular and we understand $\mathbf{A}_{\mathrm{FE}}^{-1}$ as a mapping $\mathbf{A}_{\mathrm{FE}}^{-1} : V' \to V$.

We propose the preconditioned conjugate gradient (pcg) method for the solution of this linear system. The pcg algorithm constructs a sequence $(\mathbf{u}^i)_{i\in\mathbb{N}}$ of grid functions $\mathbf{u}^i \in V$ that converges to the exact solution of (2.4). For completeness, we recall the algorithmic formulation of the pcg method. We denote by $\langle\cdot,\cdot\rangle$ the Euclidean scalar product and by $\|\cdot\|$ the corresponding norm. The algorithm will be terminated if the Euclidean norm of the residual divided by the Euclidean norm of the right-hand side is smaller than a prescribed stopping criterion $\varepsilon > 0$.

**Algorithm 4.1 (pcg algorithm)**

***Choose a starting guess*** $\mathbf{u}^0 \in V$ *(e.g.,* $\mathbf{u}^0 = (0,0,\dots 0)^{\mathsf{T}}$*);*

***let***

$$\mathbf{r}^0 := \mathbf{F} - \mathbf{A}\mathbf{u}^0; \qquad \mathbf{s}^0 := \mathbf{A}_{\mathrm{FE}}\mathbf{r}^0; \qquad \mathbf{p}^0 := \mathbf{s}^0;$$

***for*** $i = 0, 1, 2, \dots$ ***do***

***if*** $\|\mathbf{r}^i\| \,/\, \|\mathbf{F}\| \geq \varepsilon$ ***then begin***

$$\begin{aligned}
\alpha_{i+1} &:= \frac{\langle \mathbf{r}^i, \mathbf{p}^i\rangle}{\langle \mathbf{A}\mathbf{p}^i, \mathbf{p}^i\rangle}\\
\mathbf{u}^{i+1} &:= \mathbf{u}^i + \alpha_{i+1}\mathbf{p}^i;\\
\mathbf{r}^{i+1} &:= \mathbf{r}^i - \alpha_{i+1}\,\mathbf{A}\mathbf{p}^i\,;\\
\mathbf{s}^{i+1} &:= \mathbf{A}_{\mathrm{FE}}^{-1}\mathbf{r}^{i+1};\\
\beta_{i+1} &:= \frac{\langle \mathbf{s}^{i+1}, \mathbf{r}^{i+1}\rangle}{\langle \mathbf{s}^i, \mathbf{r}^i\rangle};\\
\mathbf{p}^{i+1} &:= \mathbf{s}^{i+1} + \beta_{i+1}\mathbf{p}^i;
\end{aligned} \tag{4.1}$$

***end else "the approximate solution is*** $\mathbf{u}^i$***"; stop;***

***end;***

The convergence rate of the pcg method is given in the next theorem. The error will be measured in the "energy norm" which is defined by $\|\mathbf{u}\|_{\mathbf{A}} := \langle \mathbf{u}, \mathbf{A}\mathbf{u} \rangle^{1/2}$.

**Theorem 4.2** *Let the spectrum* $\sigma\left(\mathbf{A}_{\mathrm{FE}}^{-1}\mathbf{A}\right)$ *satisfies*

$$\sigma(\mathbf{A}_{\mathrm{FE}}^{-1}\mathbf{A}) \subseteq [a, b], \quad \kappa := b/a \geq 1\,.$$

*Then for all* $\mathbf{u}^0 \in V$ *and* $i \geq 1$

$$\|\mathbf{u}^i - \mathbf{u}\|_{\mathbf{A}} \leq \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^i \|\mathbf{u}^0 - \mathbf{u}\|_{\mathbf{A}}\,. \tag{4.2}$$

**Remark 4.3** *For the recovery method, the convergence rate can be estimated by*

$$\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \leq \frac{\sqrt{\delta_{\max}\overline{n}} - 1}{\sqrt{\delta_{\max}\overline{n}} + 1}.$$

Note that in every iteration step a system of linear equations of the form

$$\mathbf{A}_{\mathrm{FE}}\mathbf{v} = \mathbf{g}$$

has to be solved. In contrast to the original equation the matrix $\mathbf{A}_{\mathrm{FE}}$ stems from a finite element discretization of a continuous Poisson-type PDE. Hence, efficient multigrid solvers are available to solve this system in linear complexity (cf., e.g., [1], [12], [22], [17], [5], [21], [3], [28]). This leads to a nested iteration, where the outer iteration is given by the pcg method while the inner iteration replaces the step (4.1) by a multigrid solver. We do not describe the multigrid method for elliptic PDEs for problems with discontinuous coefficients in detail here but refer to literature instead (cf., e.g., [22]).

# 5 Dirichlet-type Constraints

In this section, we will consider the problem where the values of the solution have prescribed value zero on a subset $\Theta_D$ with $\emptyset \neq \Theta_D \subsetneq \Theta$.

**Definition 5.1** *For a prescribed subset* $\omega \subset \Theta$ *the space* $S(\omega)$ *is*

$$S(\omega) := \left\{\mathbf{u} \in \mathbb{R}^{\Theta} \mid \forall x \in \omega : \mathbf{u}(x) = 0\right\}$$

Hence, for the prescribed subset $\Theta_D$ the space of grid functions is given by $S(\Theta_D)$.

The recovery method for the derivation of the bilinear form $B_{\mathrm{FE}}$ is applied verbatim as for the unconstrained problem (2.3) and the definition (3.15), (3.17) is used without changes. However, the finite element space $S^{\mathrm{FE}}$ (cf. (3.16)) has to take into account the essential constraints and we set

$$S_D^{\mathrm{FE}} := \left\{u \in S^{\mathrm{FE}} \mid \forall x \in \Theta_D : u(x) = 0\right\}. \tag{5.1}$$

**Remark 5.2** *(a) Note that the evaluation of functions $u \in H^1(\Omega)$ at discrete points $x \in \Omega$ is not defined in general since $H^1(\Omega) \not\subset C^0(\Omega)$. However, for finite element functions $u \in S^{\mathrm{FE}}$, the point evaluation is well defined.*

*(b) The recovery method can be interpreted as the inverse of the transfer "boundary value problem and basis of the finite element space→stiffness matrix" in the following sense. Consider the special case $\mathcal{G} = \mathcal{G}_{\mathrm{FE}}$, where $\mathcal{G}_{\mathrm{FE}}$ is a triangulation of a domain $\Omega$ with boundary $\Gamma$. Assume the lattice equations originates from the finite element discretisation of the continuous Laplace problem with homogeneous Dirichlet boundary conditions at $\Gamma$ on the mesh $\mathcal{G}_{\mathrm{FE}}$. Then, the usual finite element space on $\mathcal{G}_{\mathrm{FE}}$ for the Dirichlet problem coincides with the recovered space $S_D^{\mathrm{FE}}$ as in (5.1).*

The proof that the bilinear form for the lattice equation and the bilinear form on the continuous level have equivalent energies is a repetition of the proof of Theorem 3.14. The constants of equivalency are the same as in Theorem 3.14.

Similarly, Theorem 4.2 holds verbatim for the problem with Dirichlet constraints.

# 6 Numerical Experiments

The following constants enter the convergence estimates of the pcg method (cf. Remark 4.3).

- The maximal "overlap" of the paths

$$\overline{n} := \max \left\{ \max_{e \in \mathcal{E}} \operatorname{card} \left\{ \tilde{e} \in \mathcal{E}_{\mathrm{FE}} : e \in \pi(\tilde{e}) \right\}, \max_{\tilde{e} \in \mathcal{E}_{\mathrm{FE}}} \operatorname{card} \left\{ e \in \mathcal{E} : \tilde{e} \in \pi_{\mathrm{FE}}(e) \right\} \right\}. \tag{6.1}$$

- The maximal weighted sum of coefficients along the paths, i.e., $\delta_{\max}$ as in (3.3). Note that $\delta_{\max}$ is the average of various parameters and we will investigate the dependence on these parameters separately:

  - The maximal ratio of the lengths of segments in paths $\pi_{\mathrm{FE}}(e)$ compared to the length of $e$ and vice versa:

$$\overline{\eta} := \max \left\{ \max_{e \in \mathcal{E}} \max_{\tilde{e} \in \pi_{\mathrm{FE}}(e)} h_e / h_{\tilde{e}}, \max_{e \in \mathcal{E}} \max_{\tilde{e} \in \pi_{\mathrm{FE}}(e)} h_{\tilde{e}} / h_e \right\}, \tag{6.2}$$

  - the length of the paths

$$\overline{q} := \max \left\{ \max_{e \in \mathcal{E}} \operatorname{card} \pi_{\mathrm{FE}}(e), \max_{\tilde{e} \in \mathcal{E}_{\mathrm{FE}}} \operatorname{card} \pi(\tilde{e}) \right\}. \tag{6.3}$$

– the magnitude and the variations of the conductivity coefficients $(a_e)_{e\in\mathcal{E}}$ along the paths $\pi(\tilde{e})$ and $\pi_{\mathrm{FE}}(e)$.

The goal of the numerical experiments is to investigate the performance of the recovery method and the sharpness of the convergence estimates by performing the following experiments. The right-hand side for problem (2.4) is chosen by

$$F_x := \sin(x_1) + e^{x_2} - c \qquad \forall x = (x_1, x_2) \in \Theta,$$

where $c$ is the constant such that $\sum_{x\in\Theta} F_x = 0$. For this problem, the pcg algorithm (Algorithm 4.1) is chosen as the linear solver with stopping criterion $\varepsilon = 10^{-8}$. We emphasize that the performance of the idealized pcg method is investigated where in step (4.1) the exact solution of $\mathbf{A}_{\mathrm{FE}}\mathbf{s}^{i+1} = \mathbf{r}^{i+1}$ is employed. In practical applications this step has to be replaced by a fast PDE solver such as multigrid. Since we are interested in the systematic study of the effect of using $\mathbf{A}_{\mathrm{FE}}$ as a preconditioner for $\mathbf{A}$ we preferred to use the matrix $\mathbf{A}_{\mathrm{FE}}$ instead of its approximation via a multigrid solver.

From the numerical experiments averaged convergence rates $\tilde{\lambda}$ are derived as follows. Assume that the pcg method needs $m$ iterations to terminate. Then, we set

$$\tilde{\lambda} := \varepsilon^{1/m}.$$

This number expresses the averaged reduction factor of the Euclidean norm of the residuals in each iteration step. This number will be compared with the ratio $\lambda := (\sqrt{\kappa} - 1)/(\sqrt{\kappa} + 1)$ which is the averaged reduction of the energy norm of the iteration error as predicted by Theorem 4.2.

## 6.1 Dependence of the Convergence Rates on the Problem Size (Example 1 and 2)

In this subsection, we will investigate the dependence of convergence rate on the size of the problem, i.e., on $\dim V$. In order to study this behaviour independently of the other mesh constants we have, in a first experiment, specified a lattice on a reference cell and then defined a sequence of increasingly finer meshes by shrinking and periodically copying the reference cell to a larger mesh. In this case, the constants characterizing the mesh stays constant and we can investigate the effect of increasing dimensions of the problem isolatedly. We have considered two test cases: for the first one (first row in Figure 1), the structure in the master cell is very simple while the lattice in the second master cell (second row in Figure 1) is more unstructured. The conductivity coefficient $a_e$ was chosen to be 1 for all edges. The configurations in the master cells are depicted and the refined meshes at refinement level $\ell = 5$.Relevant

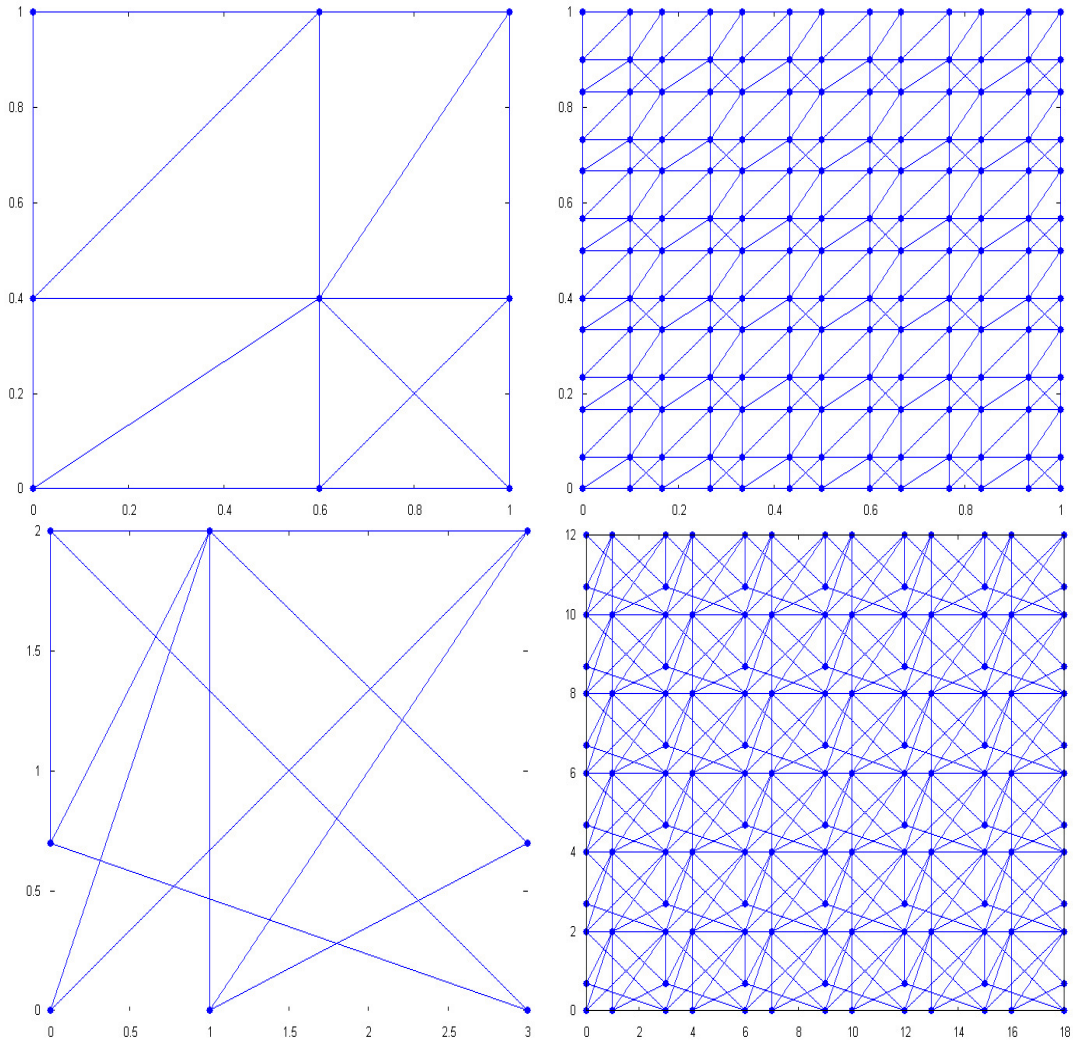


Figure 1: First row: Master cell for Example 1 and periodic refinement. Second row: Master cell for Example 2 and periodic refinement

constants for these two meshes are listed below.

| | $C_{sr}$ | $\overline{n}$ | $\overline{\eta}$ | $\overline{q}$ | $\delta_{\max}$ |
|---|---|---|---|---|---|
| Example 1 | 5.2 | 2 | 1 | 2 | 1.1 |
| Example 2 | 7.3 | 3 | 5.2 | 4 | 1.8 |

The averaged convergence rates and the number of iterations are depicted in the following table.

| ref. level | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ex. 1** | | | | | | | | | | | |
| $\tilde{\lambda}$ | 0.03 | 0.24 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.32 | 0.32 | 0.32 |
| ♯ of it. | 5 | 13 | 15 | 15 | 15 | 15 | 15 | 16 | 16 | 16 | 16 |
| **Ex. 2** | | | | | | | | | | | |
| $\tilde{\lambda}$ | 0.03 | 0.27 | 0.38 | 0.43 | 0.46 | 0.46 | 0.46 | 0.48 | 0.48 | 0.48 | 0.48 |
| ♯ of it. | 5 | 14 | 19 | 22 | 24 | 24 | 24 | 25 | 25 | 25 | 26 |

The main observation is that the averaged convergence rates $\tilde{\lambda}$ stay bounded properly away from 1 with increasing dimension and range between 0.24 and 0.5.

The difference in the convergence rates for the two examples can be explained because the ratio $\overline{\eta}$ (cf. (6.2)) equals 1 for Example 1 while it is 5.15 for Example 2. The parameter $\overline{\eta}$ is defined as the maximum over local length ratios (cf. 6.2). The experiments show that, if the length ratios which characterize the maximum in $\overline{\eta}$, are distributed "uniformly" over the domain as is the case for Examples 1 and 2, the predicted qualitative dependence of the convergence rates on this quantity is visible also in the numerical experiments.

However, in both cases the convergence rates are properly bounded away from 1 and the numbers of iterations to reach the stopping criterion are quite moderate.

Note that lattices which arise by periodically copying lattice configuration on reference cells are well suited for systematic testing the sensitivity of the algorithm for larger dimensions while in practical applications periodic lattices are rather exceptional. More typical are applications such as the "routing channel" (cf. Figure 2), where the dimension of the linear system is large while the geometry contains no systematic periodicity. We will address the performance of the recovery for this example in Subsection 6.5.

## 6.2 Dependence on the Path Lengths (Example 3)

In this subsection we will investigate systematically the dependence of the recovery method on the maximal lengths of the paths, i.e., on the quantity $\overline{q}$ as defined in (6.3). The lattice topology is as for Example 1 with the exception that an additional edge $e$ is inserted which connects the left bottom corner node with the right top corner (cf. Figure 3). With increasing refinement level $\ell$ the path length $\pi_{\mathrm{FE}}(e)$ for this exceptional edge becomes increasingly large.The relevant constants are listed below

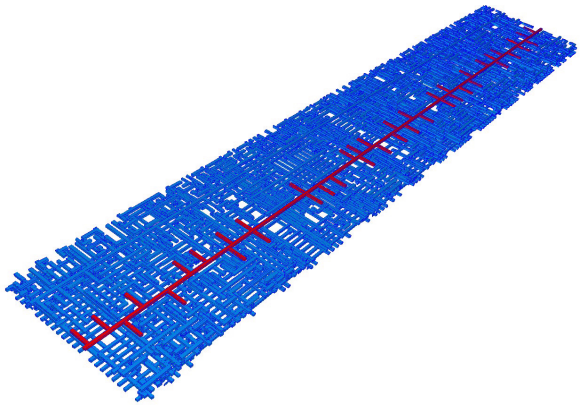

Figure 2: Picture of the routing channel.

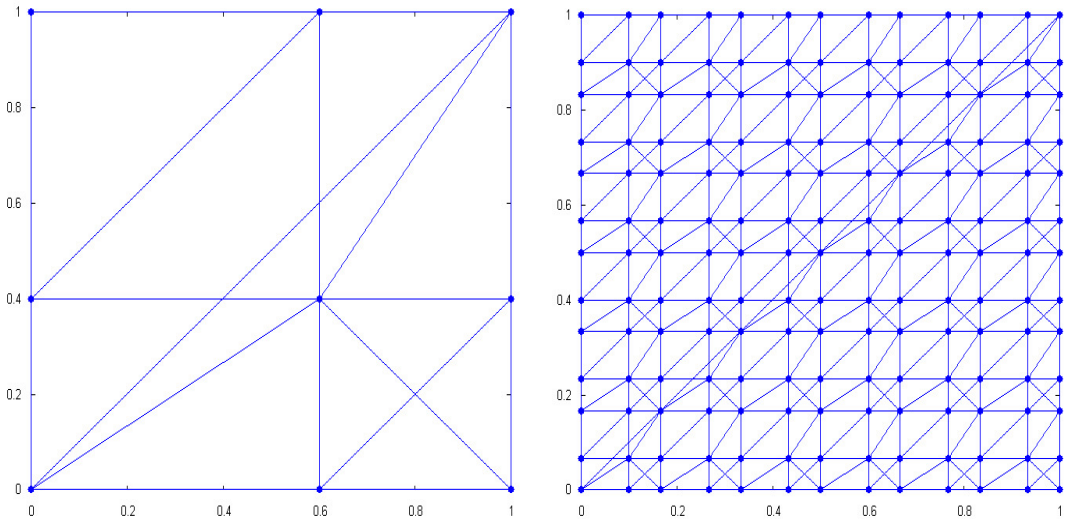


Figure 3: Master cell for Example 3 and periodic refinement.

| | $C_{sr}$ | $\overline{n}$ | $\overline{\eta}$ | $\overline{q}$ | $\delta_{\max}$ |
|---|---|---|---|---|---|
| $\ell = 0$ | 5.2 | 4 | 1 | 3 | 1.8 |
| $\ell = 1$ | 5.2 | 8 | 1 | 3 | 1.3 |
| $\ell = 2$ | 5.2 | 12 | 1 | 3 | 1.2 |
| $\ell = 3$ | 5.2 | 16 | 1 | 4 | 1.1 |
| $\ell = 4$ | 5.2 | 20 | 1 | 4 | 1.1 |
| $\ell = 5$ | 5.2 | 24 | 1 | 4 | 1.1 |

The numerical results are depicted in the following table.

| ref. level | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $\widetilde{\lambda}$ | 0.03 | 0.24 | 0.27 | 0.29 | 0.29 | 0.32 |
| ♯ of it. | 5 | 13 | 14 | 15 | 15 | 16 |

We see that the dependence on the parameters $\overline{n}$ and $\overline{q}$ is harmless for this example and behaves better as predicted by theory. The reason is that the length of only *one* path in the lattice (connecting the left bottom nodal point with the right top nodal point) is increased. However, we expect that for "pathological" examples where *all* paths are very long the theoretical estimates are sharp while for meshes with only few long paths the performance of the method is not significantly influenced.

## 6.3 Dependence on the Maximal "Overlap" of the Paths, the Lengths of the Paths, and the Shape-Regularity Constant (Example 4)

In this subsection, we will investigate the dependence of the recovery method on a geometry where the mesh constants systematically become degenerate. We have depicted the lattice geometry for the refinement levels $\ell = 0$ and $\ell = 5$. The minimal angles of the triangles at the top of the rectangle (down to the horizontal line in the middle) tend to zero with increasing refinement level. The length of the path in the Delaunay mesh connecting the points $(1, 0)$ and $(0.2, 0.3)$ also increases with increasing refinement level.

| | $C_{sr}$ | $\overline{n}$ | $\overline{\eta}$ | $\overline{q}$ | $\delta_{\max}$ |
|---|---|---|---|---|---|
| $\ell = 2$ | 11.1 | 2 | 1.0 | 2 | 1.0 |
| $\ell = 4$ | 15.3 | 2 | 1.3 | 3 | 1.0 |
| $\ell = 6$ | 19.7 | 4 | 1.1 | 5 | 1.0 |
| $\ell = 8$ | 25.1 | 5 | 1.1 | 6 | 1.0 |
| $\ell = 10$ | 30.7 | 6 | 1.0 | 7 | 1.0 |
| $\ell = 12$ | 36.4 | 7 | 1.0 | 8 | 1.0 |
| $\ell = 14$ | 42.0 | 7 | 1.6 | 8 | 1.0 |
| $\ell = 16$ | 50.5 | 9 | 1.0 | 10 | 1.0 |
| $\ell = 18$ | 53.4 | 10 | 1.0 | 11 | 1.0 |
| $\ell = 20$ | 59.1 | 10 | 1.7 | 11 | 1.0 |

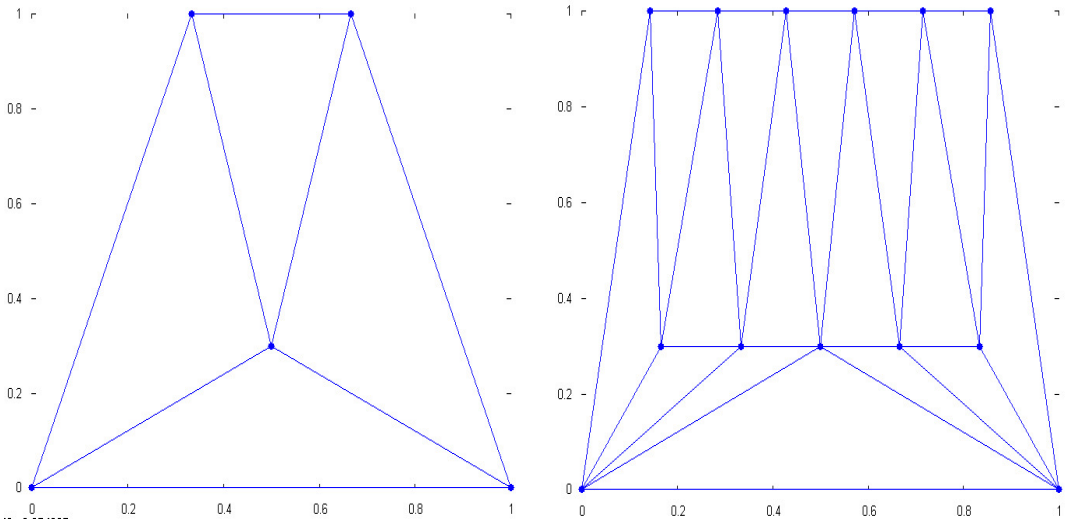


Figure 4: Master cell and periodic refinement for Example 4.

The convergence rates and the number of iterations are listed below.

| ref. level | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\tilde{\lambda}$ | 0.0 | 0.01 | 0.05 | 0.07 | 0.10 | 0.10 | 0.10 | 0.13 | 0.13 | 0.13 |
| ♯ of it. | 1 | 4 | 6 | 7 | 8 | 8 | 8 | 9 | 9 | 9 |

We see that the dependence of the recovery method on these parameters is very moderate and much better as predicted by theory. The reason is that the number of edges which lead to the large overlap constant is small compared to the number of edges with small number of overlapping paths. Again, we expect that there are "pathological" lattices where the estimates in our theory become sharp while for practical problems we expect that this dependence will not have a big influence.

## 6.4 Dependence on the Magnitude and Variations of the Conductivity Coefficients (Examples 5-8)

In this subsection, we investigate the dependence of the recovery method on the size and the variations of the conductivity coefficients on the given mesh. As test problems we have again specified a lattice configuration on a reference cell and obtained the final lattice by shrinking the cell and periodically copying the reference cell.

The different geometries (including also some degenerate ones) on the reference cell and the final lattice are depicted in Figure 5.We have considered the following

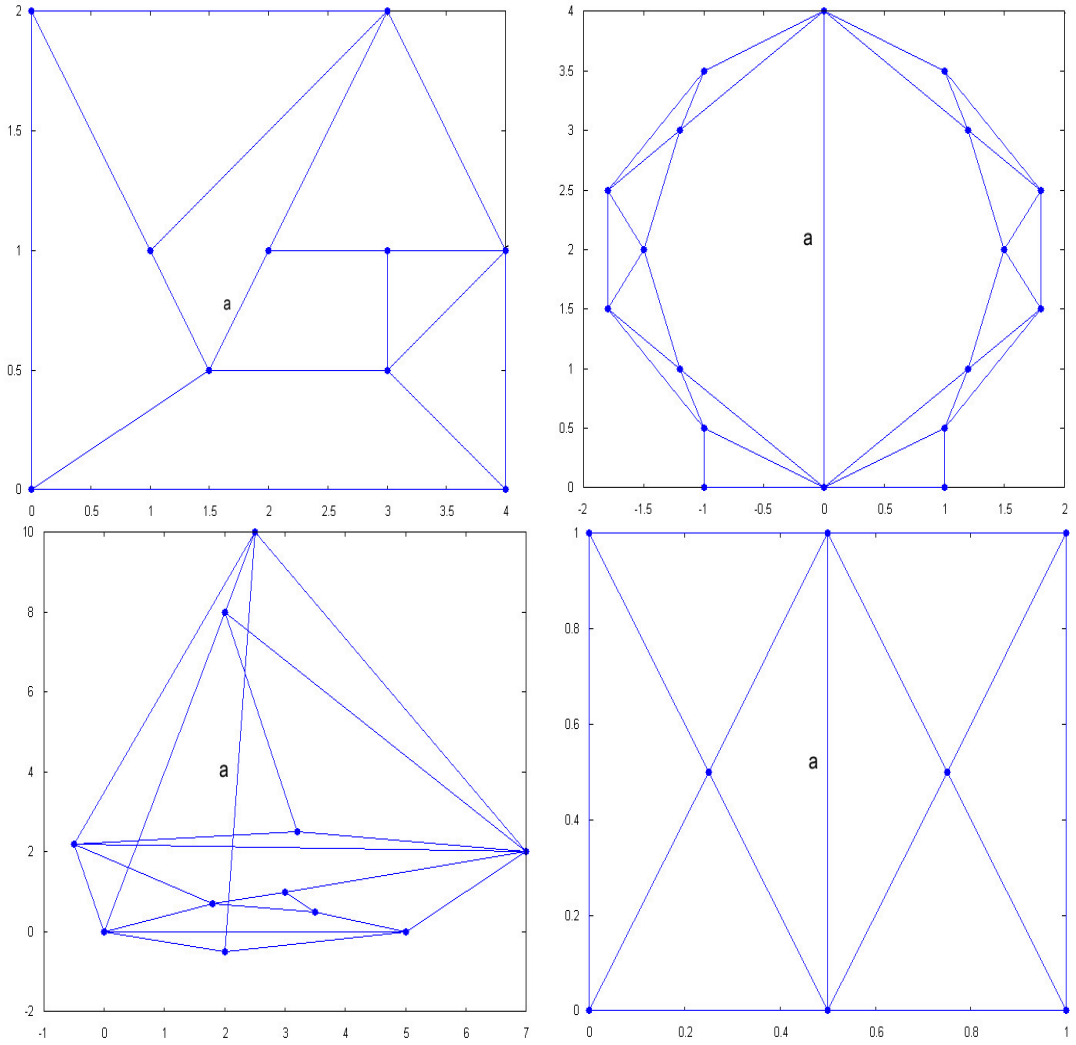


Figure 5: Top left: Example 5, Top right: Example 6, Bottom left: Example 7, Bottom right: Example 8.

configurations for the conductivity coefficient $(a_e)_{e\in\mathcal{E}}$

| | |
|---|---|
| Configuration 0: | (Reference configuration) All coefficients are set to 1. |
| Configuration 1a: | Values 1 and 100 are distributed on the edges in an alternating way, |
| Configuration 1b: | Values 1 and $\frac{1}{100}$ are distributed on the edges in an alternating way, |
| Configuration 2a: | Values 1 and $10^6$ are distributed on the edges in an alternating way, |
| Configuration 2b: | Values 1 and $10^{-6}$ are distributed on the edges in an alternating way, |
| Configuration 3: | The edge marked by "a" in the picture is set to $10^6$ and the others to 1, |

The relevant constants are listed below.

| | $C_{sr}$ | $\overline{n}$ | $\overline{\eta}$ | $\overline{q}$ | $\delta_{\max}$ | | $C_{sr}$ | $\overline{n}$ | $\overline{\eta}$ | $\overline{q}$ | $\delta_{\max}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ex. 5** | | | | | | **Ex. 6** | | | | | |
| Config. 0 | 17.6 | 2 | 1.4 | 3 | 1.0 | Config. 0 | 16.6 | 4 | 1.0 | 4 | 1.0 |
| Config. 1a | 17.6 | 2 | 1.4 | 3 | 1.0 | Config. 1a | 16.6 | 4 | 1.0 | 4 | 1.0 |
| Config. 1b | 17.6 | 2 | 1.6 | 2 | 1.0 | Config. 1b | 16.6 | 4 | 1.0 | 4 | 2.1 |
| Config. 2a | 17.6 | 2 | 1.4 | 3 | 1.0 | Config. 2a | 16.6 | 4 | 1.0 | 4 | 1.0 |
| Config. 2b | 17.6 | 2 | 1.6 | 2 | 1.0 | Config. 2b | 16.6 | 4 | 1.0 | 4 | 11158.5 |
| Config. 3 | 17.6 | 2 | 1.4 | 3 | 1.0 | Config. 3 | 16.6 | 4 | 1.0 | 4 | 86764.1 |
| **Ex. 7** | | | | | | **Ex. 8** | | | | | |
| Config. 0 | 66.6 | 3 | 2.7 | 4 | 1.1 | Config. 0 | 8.5 | 2 | 1.1 | 2 | 1.0 |
| Config. 1a | 66.6 | 3 | 2.7 | 4 | 14.2 | Config. 1a | 8.5 | 2 | 1.1 | 2 | 1.0 |
| Config. 1b | 66.6 | 3 | 2.7 | 3 | 6.9 | Config. 1b | 8.5 | 3 | 2.7 | 3 | 4.7 |
| Config. 2a | 66.6 | 3 | 2.7 | 4 | 132600.8 | Config. 2a | 8.5 | 2 | 1.1 | 2 | 1.0 |
| Config. 2b | 66.6 | 3 | 2.7 | 3 | 60496.3 | Config. 2b | 8.5 | 2 | 1.1 | 2 | 37268.8 |
| Config. 3 | 66.6 | 3 | 2.7 | 4 | 54807.7 | Config. 3 | 8.5 | 2 | 1.1 | 2 | 149073.1 |

The corresponding convergence rates and numbers of iterations are listed in the following table.

| | $\check{\lambda}$ | ♯ of it. | | $\check{\lambda}$ | ♯ of it. |
|---|---|---|---|---|---|
| **Ex. 5** | | | **Ex. 6** | | |
| Config. 0 | 0.05 | 6 | Config. 0 | 0.1 | 8 |
| Config. 1a | 0.05 | 6 | Config. 1a | 0.05 | 6 |
| Config. 1b | 0.01 | 6 | Config. 1b | 0.03 | 5 |
| Config. 2a | 0.01 | 4 | Config. 2a | 0.01 | 4 |
| Config. 2b | 0.01 | 4 | Config. 2b | 0.002 | 3 |
| Config. 3 | 0.03 | 5 | Config. 3 | 0.16 | 10 |
| **Ex.7** | | | **Ex. 8** | | |
| Config. 0 | 0.13 | 9 | Config. 0 | 0.002 | 3 |
| Config. 1a | 0.13 | 9 | Config. 1a | 0.002 | 3 |
| Config. 1b | 0.1 | 8 | Config. 1b | 0.1 | 8 |
| Config. 2a | 0.16 | 10 | Config. 2a | 0.0001 | 2 |
| Config. 2b | 0.07 | 7 | Config. 2b | 0.27 | 14 |
| Config. 3 | 0.21 | 12 | Config. 3 | 0.01 | 4 |

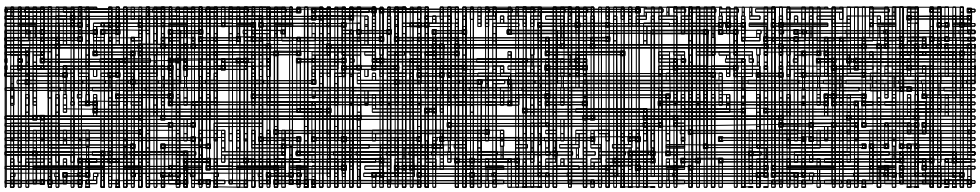

Figure 6: Graph of the routing channel.

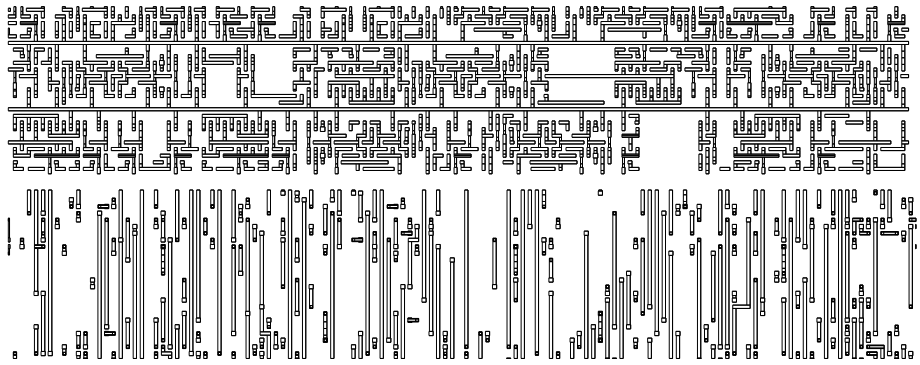

Figure 7: Two subgraphs in the routing channel.

From these tables, it is clearly visible that the theoretically predicted (negative) influence of jumping coefficients on the convergence rates is very moderate. In most test cases the convergence rates are even improved compared to the case of constant coefficient.

## 6.5 Application to Routing Channel

In this subsection, we will report the numerical results for the real-life problem of a *routing channel* and relate these results to the systematic parameter studies as in the previous sections. The graph is depicted in Figure 6. This graph is unstructured and strongly anisotropic (with respect to the lengths of edges in the graph). The lengths of the edges vary by four orders of magnitude. This is illustrated In Figure 7, where two subgraphs in this lattice are depicted.

The lattice contains 14042 nodes and 20366 edges. In view of the strong anisotropies it is clear that the Delaunay triangulation for this set of mesh points contains triangles with very small angles. The Delaunay mesh is depicted in Figure 8. The quality constants of the mesh are as follows.

| $C_{sr}$ | $\overline{n}$ | $\overline{\eta}$ | $\overline{q}$ | $\delta_{\max}$ |
|---|---|---|---|---|
| 266.7 | 4578 | 2745.3 | 137 | $3.51 \times 10^7$ |

Hence, from our theory we expect that the efficiency of the pcg-algorithm is significantly reduced. This is underpinned by the numerical experiment where 4753

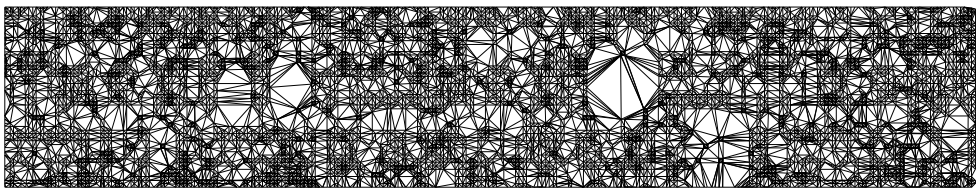

Figure 8: Delaunay mesh for the routing channel.

iterations are needed in order to reduce the residual below $10^{-8}$. Future research will be directed towards a refinement of this algorithm to handle strongly anisotropic edge lengths in the lattice.

**Acknowledgements:** We thank T. Meienberg and A. Veit for the implementation of the recovery algorithm. A. Veit also carried out the numerical experiments.

Thanks are due to Dr. C. Lage who supplied us with the data of the routing channel.

The results have been achieved during stays of the second author at the Institute for Computational Engineering and Sciences (ICES) at the University of Texas. This support is greatly acknowledged.

# References

[1] R. Alcouffe, A. Brandt, J. Dendy, and J. Painter. The multi-grid method for the diffusion equation with strongly discontinuous coefficients. *SIAM J. Sci. Stat. Comput.*, 2(4):430–454, 1981.

[2] I. Babuška and S. Sauter. Efficient Solution of Lattice Equations by the Recovery Method. Part 1: Scalar Elliptic Problems. *Comp. Vis. Sci.*, 7(3-4):113–119, 2004.

[3] D. Braess. Towards Algebraic Multigrid for Elliptic Problems of Second Order. *Computing*, 55(4):379–393, 1995.

[4] J. Bramble. *Multigrid Methods*. Pitman Research Notes in Mathematics. Longman Scientific & Technical, 1993.

[5] A. Brandt. Algebraic Multigrid Theory: The symmetric case. *Appl. Math. Comput.*, 19:23–56, 1986.

[6] Z. Chen and T. Hou. A mixed multiscale finite element method for elliptic problems with osciallating coefficients. *Math.Comp.*, 72:541–576, 2003.

[7] L. P. Chew. Guaranteed-quality Delaunay meshing in 3D. In *Proc. 13th Symp. Comp. Geom.*, pages 391–393. ACM, 1997.

[8] G. Constantinides and A. C. Payatakes. A Three Dimensional Network Model for Consolidated Porous Media. Basic Studies. *Chem. Eng. Comm.*, 81:55–81, 1980.

[9] I. Fatt. The Network Model of Porous Media. *Trans. Am. Inst. Mem. Metall Pet. Eng.*, 207:144–181, 1956.

[10] P. George. *Automatic Mesh Generation and Finite Element Computation*, volume IV, chapter Finite Element Methods (Part 2), pages 69–192. North-Holland, 1996. In: Handbook of Numerical Analysis, Eds.: P.G. Ciarlet and J.L. Lions.

[11] L. Gibson and M. Ashby. *Cellular Solids, Structures and Properties.* Pergamon Press, Exeter, 1989.

[12] M. Griebel and S. Knapek. A multigrid-homogenization method. In W. Hackbusch and G. Wittum, editors, *Modeling and Computation in Environmental Sciences*, pages 187–202, Braunschweig, 1997. Vieweg. Notes Numer. Fluid Mech. 59.

[13] W. Hackbusch. *Multi-Grid Methods and Applications.* Springer Verlag, Berlin, 1985, 2nd edition 2003.

[14] J. C. Hansen, S. Chien, R. Skalog, and A. Hoger. A Classic Network Model Based on the Structure of the Red Blood Cell Membrane. *Biophy. J.*, 70:146–166, 1996.

[15] T. Hou and X. Wu. A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media. *J. Comput. Phys.*, 134:169–189, 1997.

[16] T. Y. Hou, X. H. Wu, and Z. Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp.*, 68:913–943, 1999.

[17] J. Mandel, M. Brezina, and P. Vaněk. Energy optimization of algebraic multigrid bases. *Computing*, 62:205–228, 1999.

[18] M. Ostoja-Starzewski. Lattice Models in Micromechanics. *App. Mech. Review*, 55(1):35–60, 2002.

[19] M. Ostoja-Starzewski, P. Y. Shang, and K. Alzebdoh. Spring Network Models in Elasticity and Fractures of Composites and Polycristals. *Comp. Math. Sci.*, 7:89–93, 1996.

[20] G. I. Pshenichnov. *Theory of Lattice Plates and Shells.* World Scientific, Singapure, 1993.

[21] J. Ruge and K. Stüben. Algebraic multigrid. In S. McCormick, editor, *Multigrid Methods*, pages 73–130, Pennsylvania, 1987. SIAM Philadelphia.

[22] S. Sauter and R. Warnke. Composite Finite Elements for Elliptic Boundary Value Problems with Discontinuous Coefficients. *Computing*, 77(1):29–55, 2006.

[23] J. Shewchuk. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In *First Workshop on Appl. Geom.*, pages 124–133, Philadelphia, Pennsylvania, USA, 1996. Association for Computing Machinery.

[24] J. Shewchuk. Tetrahedral Mesh Generation by Delaunay Refinement. In *Proc. Of the 14th Annual Symp. On Comp. Geom.*, pages 86–95, Minneapolis, USA, 1998. Association of Computing Machinery.

[25] S. Shu, I. Babuška, J. Xu, Y. Xiao, and L. Zikatanov. Preconditioning Discrete Models of Lattice Block Materials. Technical report, Penn State University, 2007.

[26] U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, London, 2001.

[27] J. Xu. Iterative methods by space decomposition and subspace corrections. *SIAM Rev.*, 34:581–613, 1992.

[28] J. Xu and L. Zikatanov. On an energy minimization basis in algebraic multigrid methods. *Comp. Vis. Sci.*, 7(3-4):121–127, 2004.